4-1-2009

Self-Assessments by U.S. Army Officers: Effects of Skill Level and Item Ambiguity on Accuracy

John T. Breidert
Western Kentucky University, john.breidert781@wku.edu

Follow this and additional works at: http://digitalcommons.wku.edu/theses

 Part of the Industrial and Organizational Psychology Commons

SELF-ASSESSMENTS BY U.S. ARMY OFFICERS:
EFFECTS OF SKILL LEVEL AND ITEM AMBIGUITY ON ACCURACY

A Thesis
Presented to
The Faculty of the Department of Psychology
Western Kentucky University
Bowling Green Kentucky

In Partial Fulfillment
Of the Requirements for the Degree
Master of Industrial/Organizational Psychology

By
John T. Breidert

May 2009

SELF-ASSESSMENTS BY U.S. ARMY OFFICERS:
EFFECTS OF SKILL LEVEL AND ITEM AMBIGUITY ON ACCURACY

Date Recommended ___April 29, 2009___

Elizabeth L. Shoenfelt_____
Director of Thesis

Reagan D. Brown_____

Scott B. Shadrick_____

_____
Dean, Graduate Studies and Research   Date

Acknowledgements

I would like to thank first of all the U. S. Army and the officers that were involved in participating in this thesis effort. I would also like to thank everyone at Army Research Institute for the Behavioral and Social Sciences at Fort. Knox, KY. This includes Dr. Scott Shadrick, Dr. Jeffrey Fite, and Dr. Christopher Vowels to name a few. Thanks to my thesis chair, Dr. Betsy Shoenfelt, for her good advice and quick turnaround during the editing phases. My family and Kari, my adorable lady-friend, have been quite helpful in allowing me to vent steam when necessary. Finally, I would like to thank all of my compatriots here, the 2009 I/O Psychology class. Without their ability to play pong and socialize, I would never have made it.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

Page

SELF-ASSESSMENTS BY U.S. ARMY OFFICERS:
EFFECTS OF SKILL LEVEL AND ITEM AMBIGUITY ON ACCURACY

John T. Breidert   May 2009   62 Pages

Directed by: Elizabeth L. Shoenfelt, Ph.D, Scott B. Shadrick, Ph.D, and
Reagan D. Brown, Ph.D

Department of Psychology   Western Kentucky University

Abstract

Organizations benefit from the use of training and performance assessments. Self-assessment is a way for trainees to monitor their progress throughout training and on the job. The literature indicates that ambiguity and skill level are factors that impact the accuracy of self-assessments. Previously, the effect of the interaction of ambiguity and skill level on self-assessment accuracy had not been investigated. The present study assessed the effect of skill level and item ambiguity on the accuracy of self-assessments made by Lieutenants and Captains in the U. S. Army. The results indicated that increased skill level resulted in increased accuracy of self-assessments while ambiguity had no effect. Counter to the hypothesis, as items became more ambiguous, both Captains and Lieutenants self-assessed more accurately. Implications and limitations are discussed, along with recommendations for future research.

Introduction and Review of the Literature

Many terms are used to describe the process of self-assessment. One could identify several such as self-evaluation, self-grading, self-appraisal, confidence ratings, and ability judgments, to name a few. Regardless of what term is used, self-assessment typically refers to an individual's evaluation of his/her own ability or standing on a given construct. The disciplines most commonly attempting self-assessment research include the health, education, business, psychology, and military fields. The U.S. Army has used self-assessment as an indication of trainee progression through distance learning training packages (Shadrick & Shaefer, 2007) and self-assessment based mini-after action review methodology (Mirabella & Love, 1998). Self-assessment can be used to monitor skill acquisition progress or necessity for further training, and can be used post-training to monitor for shifts in performance or ability. Across the spectrum of self-assessment research, the topic of most debate concerns the accuracy with which individuals can evaluate their own performance. Although some research has found self-assessment to be an accurate measure of one's standing on a given construct, the majority of research typically found it to be an under- or over-estimation of one's actual performance. This paper investigates the effects of skill level and item ambiguity as it relates to self-assessment accuracy. Following a review of the literature, a study examines the main effects of skill level and item ambiguity on U. S. Army Officers' self-assessment accuracy.

Although the accuracy of self-assessment has seen much skepticism, its benefits have been well documented. Strong, Davis, and Hawks (2004) conducted a case study involving college students enrolled in two general education classes. The benefits of

being allowed to self-assess their own grades were striking. Among those benefits were that the majority of students were more motivated to learn, understood material better, and the class was viewed as more enjoyable. Ulmer (2000) emphasized the critical thinking stemmed from reflective thinking as a major determinant for transfer of knowledge. The value of self-assessment rests upon the idea that improvement can be actively pursued once an individual recognizes their own weaknesses (e.g., Dunning, Heath, & Suls, 2004; Kruger & Dunning, 1999, 2002; Metcalfe, 1998; Strong et al., 2004; Zakay & Glicksohn, 1992). If one is able to identify one's own weaknesses, reliance upon superiors to provide feedback becomes unnecessary. Research shows that feedback is an integral part of effective training and motivation (e.g., Arnold, 1976; Koka & Hein, 2003; Locke & Latham, 1990; Shoenfelt, 1996). In the applied military setting, feedback that is immediate and accurate should translate into better military leader performance, which will have an impact on unit performance and mission success. Providing accurate feedback starts with planned observation and accurate assessment (Reider, 2008). If an individual can provide feedback to the self at all times, weaknesses are identified promptly and a trajectory toward resolution is set. Self-assessment during training can be of particular beneficence because it not only teaches the trainee to evaluate his/her own performance, it also frees the trainers/supervisors from the duty of evaluation.

Both ambiguity of self-assessment criteria and skill level of assessors are variables linked to the causes of overestimation of knowledge, skills, abilities, or performance. The ambiguity of rating-items and/or criteria is the first variable discussed. The more specific and measurable the domain and criteria are, the more accurate the self-

assessment is likely to be (e.g., Ackerman, Beier, & Bowen, 2002; Dunning, Meyerowitz, & Holzberg, 1989; Hayes & Dunning, 1997; Story, 2003; Strong et al., 2004). The second variable determined by some (e.g., Davis et al., 2006; Parker, Alford, & Passmore, 2004; Randal, Ferguson, & Patterson, 2000; Zakay & Glicksohn, 1992) to affect the accuracy of self-assessment is skill level of the participant. If the participants are experts, they possess the ability to self-assess accurately because they are familiar enough with the domain to discern differences that novices would miss. These variables are important for the effective utilization and training of self-assessment, because if not taken into account, we assume that self-assessment should be considered an inaccurate form of assessment at the macro level. Although the trend is to believe self-assessment is flawed, some contrary evidence supports its possible accuracy.

Self-Assessment is Accurate

It is important to note than in none of the studies discussed in this paper are the results of self-assessment reward contingent. Reward contingent self-assessments are likely to bias results more directly than self-assessments used merely for training or self-knowledge. A few studies contend that there is nothing wrong with self-assessment in its current state. An example of this is a study by Sullivan and Hall (1997). In an educational setting, they observed very good agreement ($r = .72$, $p < .01$) between lecturers and students on estimates of student grades. Students overestimated slightly more than underestimated their own grades. Seventy-seven percent of the self-assessments made by students in the study were within one grade level of the teacher's assessment. Only the students having the highest teacher assessments were likely to underestimate their own grades. The authors suggested that this effect could be due to

regression to the mean. Matthews and Beal (2002) found the Mission Awareness Rating Scale, a self-assessment for situation awareness, to "show promise of applicability to assessing situation awareness in field settings." Fox and Dinur (1988), in a study of males participating in a military course, found that self-assessments were significantly related to commander and peer ratings. The experimental group was told that their scores would be compared with other data to test if this would increase accuracy. Although the experimental group did not rate with any statistically significant improvement, predictive and convergent validity were found for course success, commander ratings, and peer ratings. For all but one assessment (i.e., commanders' assessments of efficiency under pressure, $p < .01$) differences between groups were insignificant ($p's > .05$). Another positive finding was that there was less of a halo effect for the self-assessments than for the peer ratings. This literature presents an argument for the capacity to effectively and accurately utilize self-assessment, while other studies contend that self-assessment produces under-estimation.

Self-Assessment Underestimates

A minority of studies has found that self-assessments *underestimate* performance. Chur-Hansen (2000) found that students in medical school evaluated themselves more severely than the tutors grading them. Another study found that trainers of educational registrars rate their own skills as lower than the ratings made by others (McKinstry, Peacock, & Blaney, 2003). Similarly, dentists rate their own work with more scrutiny than others (Milgrom, Weinstein, Ratener, Read, & Morrison, 1978). This effect may be explained by the effect of skill level and the difficulty of the task (see *Skill Level* section below). To preview, as skill level increases, accuracy increases. This is true until the

skill level is well above average, when the trainee becomes more critical and less aware of his/her own expertise. Also, as difficulty of the task increases, the likelihood of underestimating performance of the task also increases. Three circumstances have been presented here that might lead to under-estimation of self-assessment. In the next section, contrary evidence demonstrates that self-assessment is more likely to be an over-estimate of performance.

Self-Assessment Overestimates

The overwhelming majority of research reports that self-assessment overestimates true performance. Undergraduate dental students overestimated their own competence when compared to instructor marks (Mattheos, Nattestad, Falk-Nilsson, & Attstrom, 2004). Carless and Roberts-Thompson (2001) examined self-, superior-, and peer-ratings of participants in a Royal Australian Airforce training course and found that self-ratings were more lenient than ratings by others. Sidhu, Vikis, Cheifetz, and Phang (2006) found that surgeons training for laparoscopic colectomy persistently overestimated their own performance as compared to trained raters. In a study that was meant to evaluate a clinician's ability to assess their own competence using important terms in evidenced based medicine, practitioners in Sydney, Australia also overestimated their performance (Young, Glasziou, & Ward, 2002). This effect was evident even though at first glance one would assume the practitioners had self-assessed in a seemingly modest manner. Even with a low estimate of competence to begin with, no participants showed a competence that exceeded their self-rating. Only one participant's self-rating met criteria for a positive predictive value of actual competence, which contrasted the initial appearance that the practitioners were modestly rating themselves. The above are all

examples of a participant or trainee giving more credence to their own performance than indicated by the criterion measure, typically expert or supervisor ratings.

One way to describe the difference between what people think they know versus what they actually know is by use of the terms "confidence" and "competence." Confidence refers to the estimate of self-ability or skill. Confidence self-assessments can be administered pre- and post-training to evaluate an individual's feelings about their performance. Competence refers to an individual's actual standing on a given construct, and can be estimated by expert raters, or by test scores. Castle, Garton, and Kenward (2007) compared confidence scores with competence scores for nurses, doctors, and health-care assistants. The participants completed a confidence questionnaire regarding their performance in basic life support. Their competence was measured by an algorithm for basic life support that was produced by subject-matter experts (the UK Resuscitation Council). Results indicated a significant difference between confidence and competence, generally due to over-confidence. In their conclusion, the authors stated that training and exposure could increase both confidence and competence. The research regarding confidence versus competence typically shows that confidence overestimates competence. Barnsley et al. (2004) found similar results that showed an even more pronounced over-confidence effect. Their research involved Australian junior doctors and their assessments of confidence compared with actual competence for several medical competencies. For all competencies, confidence scores over-estimated actual competence. That confidence is an over-estimation of competence is another way of saying self-assessment overestimates actual performance or standing on a given construct.

Dunning et al. (2004) reported, in concurrence with the previous authors, that skill level by self-assessment is typically overrated. The researchers attributed this to the supposition that when people evaluate their own level of skill, they are overly optimistic about what they know and ignore what they do not know. This is because people are not aware of what they do not know. The process of knowing what one knows is called metacognition. It is also described as insight into one's own thinking process.   The lack of metacognitive ability leads to estimations of above-average competence.

Dunning et al. (2004) outlined four "informational deficits" that cause overestimation. The authors stated that even when people have the necessary information that would aid in accurate self-assessments, they will ignore or diminish it, leading to over-estimation of skill. Dunning et al.'s deficits are as follows:

1) Double curse of incompetence – people who are incompetent possess deficits of information that lead to errors and also block knowledge gained by these errors.

2) Unknown errors of omission – people are only cognizant of the solutions they produce, not the solutions they could have or should have produced.

3) Uncertain lessons from feedback – people assume they are adequate and inflate their perception of their own skill because feedback is often limited.

4) The ill-defined nature of competence – domains are often very general and vaguely outlined.

The fourth deficit directly relates to the next section concerning ambiguity of task statements. Dunning et al. (2004) gave us insight why people are likely to overestimate their own performance and provided another instance of literature supporting the inaccuracy and overstatement of self-assessment.

Studies evidencing both under-estimation and over-estimation show that self-assessment has a clear tendency to misevaluate performance. External factors affecting the assessment should be considered, however, before arriving at any conclusion. Item ambiguity and skill level effects must be considered. Other concerns, such as training effects, methodological issues, and individual differences, could provide additional insight into why self-assessment accuracy is so elusive. These mediating factors are ones that may influence the perception of self-assessment as an accurate form of assessment. As item ambiguity and skill level were the focus of this study, they will be addressed in more detail in the following sections.

Ambiguity

Item ambiguity refers to the level of specificity of the domain and/or items being rated by self-assessors. Evidence points toward the level of ambiguity as a determinant for how accurate self-assessments can be. Hayes and Dunning (1997) found that when the domain of possible traits is defined ambiguously, college students self-assessed more generously than their roommates rated them. Dunning et al. (1989) found that when the traits were given specific definitions, ratings tended not to be so generous. It was further noted that self-assessments showed more concurrent validity with other-ratings when the traits were well defined (Hayes & Dunning, 1997; Story, 2003). Presumably, specificity leads to decreased inflation of self-assessment by eliminating much of the uncertainty that stems from poor metacognitive ability. Specificity gives the rater a concrete referent by which objectivity takes the place of speculative subjectivity. Strong et al. (2004) suggested that to decrease the amount of inflation in assessments, thus making them more accurate, students should be given written objectives to use as a template or standard to

determine final grades. The apparent inference from these studies is that in order to improve accuracy, the domain must be clearly and specifically defined and assessed using comparative standards for self-assessment items.

Again, studies (e.g., Ackerman et al., 2002; Dunning et al., 1989; Metcalfe, 1998) show that clarity of the domain and standards of satisfactory performance aid in obtaining accurate self-assessment. Ackerman et al. (2002) stated that when assessing one's own abilities, the likelihood of under- or over-estimation is dependent upon familiarity with the ability and how broad the domain is. If the individual is unfamiliar with the ability in question, they are likely to infer from other information how they would perform. An example of this is "I could learn to synchronize visualization across relevant external players." If the domain is too broad, such as leadership, the inaccuracy of estimation is increased. The solution to this problem proposed by Ackerman et al. is to utilize more specific measures during the self-report. The authors found that broad items along with an unspecified comparison group resulted in over-estimation of ability compared with the more accurate estimation of ability when specific stimuli and absolute scales were used. The study also found significant correlations between self-reports and objective measures for the domains of science, civics, and humanities, but a non-significant correlation for the domain of business management. These findings suggest that the relationship between accuracy and ambiguity may be domain specific.

Dunning et al. (1989) described the over-confidence effect as a product of self-serving assessments of ability. The underlying theme of their four-study series is that given a certain skill, ability, or characteristic, there are usually many definitions of what constitutes a high standing or good performance. An example would be leadership

potential. If a person were asked to evaluate their own leadership potential, they would be able to conjure definitions that range from compulsive and task oriented to deliberative and people oriented, depending on the skills they themselves possess. The individual is apt to see the entire domain of behaviors possibly linked to the evaluation item and select the most self-serving combination to provide his/her rating. In the first two studies, participants produced these self-serving assessments when the trait being rated was ambiguous or subject to interpretation given a wide domain of behaviors. In the third study, it was found that as more criteria were given to produce an evaluation, the participants identified more with both positive and negative characteristics. The fourth study demonstrated that using a list of specific criteria created by an outside source tended to decrease the participants' self-serving tendencies. The problems with unclear domains and standards of performance seemed to be common issues in these studies that could be resolved by using objective domains and standards.

In similar fashion to Dunning et al. (1989), Metcalfe (1998) studied over-confidence and used the term "cognitive optimism" to explain it. According to Metcalfe, over-confidence in self-reflection is due to an individual's thought process that takes incorrect information and treats it as if it were a correct predictor of performance. Metcalfe outlined seven metacognitive phenomena that lead to over-confidence:

1) People think they will be able to solve problems when they will not.

2) People are highly confident that they are on the verge of producing the correct answer when they are, in fact, about to produce a mistake.

3) People think they know the answers to questions when they do not.

4) People think the answer is on the tip of their tongue when there is no answer, or the answer is wrong.

5) People think, even when given contradictory feedback, that they produced the correct answer and that they knew it all along.

6) People believe they have mastered learning material when they have not.

7) People think they have understood, although they are demonstrably still in the dark.

Metcalfe explained these phenomena as a product of self-deception and memory-based processing heuristics. In self-deception, people are aware on some level that their answers are or could be incorrect, but convince themselves that they are correct. Because most of the time people have no evidence to negate their correctness, most over-confidence is not seen as self-deception. The other explanation was related to memory-based processing heuristics in which people base decisions about judgments upon information retrieved from memory and information at hand that is not entirely accurate, but is treated as if it were. A person assumes that the first decision arrived upon, assembled from memory and information at hand, is correct. An example would be if one were watching Jeopardy and confidently blurted out an answer based on recollection of WWII history. The answer may have been very close because the individual had a large amount of knowledge of the subject, but he/she lacked the capability to decipher between the assumed correct answer and the actual correct answer until it was revealed.

Evidence for the heuristic explanation comes from Oskamp (1965) who studied psychiatrists and psychiatric residents that either were given a small amount of information or were given a large amount of information regarding a patient in a

hypothetical situation. The information was irrelevant to any diagnosable situation. As the amount of irrelevant information increased, so did confidence in their diagnoses. This occurred even though the participants were correct only by chance. The conclusion was that the irrelevant information was the cause of an illusion of knowledge. Ambiguity in self-assessment might allow the memory to create irrelevant information that yields a feeling of false confidence in the assessed domain. The illusion of knowledge could be minimized by using comparative standards for self-assessment that decrease ambiguity, thus allowing for confidence that matches actual performance or knowledge within the domain, and increasing accuracy.

Ambiguity was addressed by Dunning et al. (2004) when they said that one of the informational deficits that cause overestimations during self-assessment is the ill-defined nature of competence. For example, if the domain is chemistry, the solutions and knowledge are specific and either right or wrong. If the domain is essay writing, the solutions are many and the knowledge is fluid. When the domain is ill defined, people overestimate their skill; but if the domain is narrowed or specific, people are likely to estimate their skill with more accuracy.

In sum, the current state of the literature regarding ambiguity of domains and comparative standards says that the lower the ambiguity, the more likely a participant or trainee will be to give an accurate self-assessment. Familiarity with ability and domain breadth, self-serving assessments, and cognitive optimism outlined by metacognitive phenomena are all reasons given why this occurs, but the central issue at this juncture is that low ambiguity yields accurate self-assessment. Ambiguity, or any other single factor, cannot account for the entirety of self-assessment inaccuracy. However, given

proper attention to ambiguity, it could be part of a brighter future for self-assessment. Another factor that should be given proper attention is skill level, which is discussed next.

Skill Level

Studies to be discussed in this section report the accuracy of self-assessment as being mediated by skill level; people who are low in actual competency rate themselves above average. Also, people that are exceptionally gifted in some competency rate themselves lower than exceptional, but more accurately than those that are less competent. Numerous studies have found that self-assessment accuracy is determined by skill level or domain expertise. Zakay and Glicksohn (1992) found support for their hypothesis that participants who were over-confident tended to produce more wrong answers in multiple-choice tests. This was attributed to higher skilled participants being able to produce more accurate predictions of their performance. Davis et al. (2006) studied the ability of physicians to self-assess, as it has been deemed a necessary skill for identifying needed continuing education demanded by the field. A literature search that included self-assessment of ability and some performance measure for physicians found 20 studies that directly compared self-assessment with external performance measures. Analyses revealed that only seven showed positive correlations. Most notable is that many of the studies found that the worst self-assessments came from the lowest performers who actually rated themselves highest. Parker et al. (2004) used the In-training Examination (ITE) to test whether family medical residents had the capability to self-assess. The residents completed an estimation of their performance prior to being assessed with the ITE. The results showed that the residents did not predict their skills

with accuracy. Results further showed that the bottom and top quartile performers showed the biggest deficits with regard to accuracy. Top performers tended to underestimate, while bottom performers overestimated. Randal et al. (2000) tested the relationship between participants' ability to self-assess accurately and their achievement at an assessment center. They found a significant difference in the accuracy of self-assessment between those participants who were accepted (offered the job, meaning they performed better) and those that were not. Furthermore, unsuccessful participants over-estimated their performance on all exercises designed to assess competence in areas crucial to the job. The authors proposed that the successful participants possessed some ability to assess their own performance accurately, such as recognizing evaluation criteria, which allowed them to assess their own performance similarly to assessors.

An example of military research evidence in support of increased accuracy with increased skill level comes from the development and validation of the crisis response training package called Red Cape (Shadrick & Schaefer, 2007). At the beginning of training, the authors reported that the participants were prone to overconfidence (or inflated self-assessments). As the training continued, the participants reported via likert scales that the exercises were growing more challenging. This convergence from overestimation of ability to actual ability comes from training. The authors suggested that the reason for the increased awareness of actual ability in the training situation came from "increased awareness of the complexities involved in large scale interagency efforts." This finding gives support for the utilization of specific, measurable micro-components of the overall training objectives. Individual behavioral components would convey the complexity of the skill domain from the very first rating. The trainees would

become experts by examining the individual components of performance and train specifically for weak areas. Skill level in this case seemed to increase accuracy by means of increased awareness of the domain of interest.

The argument that skill level has an effect has been challenged by some to be artifactual regression toward the mean. Langendyk (2006) investigated the accuracy of third-year medical students' ability to self-assess their performance. Students assessed their own short essay, then one of their peers. The essays were then assessed by faculty. Because regression to the mean is sometimes cited to explain the phenomena of overly critical self-assessments by highly skilled individuals and overly lenient self-assessments by minimally skilled individuals, the authors compared self-grading accuracy to both peer- and faculty-graded results. Results showed that students who were rated lower by the faculty graded themselves more leniently, whereas students who were rated higher by the faculty graded themselves more critically. In fact, only students in the median skill level group graded themselves and their peers accurately according to faculty grades. The highest performing students accurately graded their peers, but were too critical of themselves. Regression to the mean was thereby ruled out as an explanation because if this phenomenon was occurring, peer grades should have been affected as much as self-grades.

Kruger and Dunning (1999, 2002) explained the relationship between skill level, (i.e., competence) and estimation of skill (i.e., confidence) using metacognition as the foundation for accuracy. As people become more competent, they are more accurately confident as opposed to over-confident. As described previously, metacognition is the process of knowing what one knows (Dunning et al., 2004). This could be extended to

include a person's ability to reflect upon one's own knowledge or behavior and to assess it continually. As one becomes more competent, one can more accurately assess performance because what constitutes good performance is now known. This effect actually leads to slight under-estimation of performance for top performers because they do not necessarily feel their performance is as good as it should be. Krueger and Mueller (2002) argued with Kruger and Dunning (1999), saying that the asymmetric errors due to differences in metacognitive ability is merely a statistical regression effect combined with the better-than-average effect. The better-than-average effect states that for a given trait, most will think they are better than 50% of the population. This is a statistical impossibility because more than half of a population cannot out-perform half of the same population. Kruger and Dunning (1999) responded, saying that the regression artifact cannot explain the entirety of the effect found in their original study, and that Krueger and Mueller used unreliable data to form their own conclusions. Because participants tended to be more accurate as their skill level increased, it seemed that there may be merit in the idea that more skilled trainees are more likely to be accurate self-assessors due to some metacognitive abilities that are produced through training.

Metacognitive ability might explain the disparity between highly skilled people and unskilled people with regard to their self-assessment accuracy. Everson and Tobias (1998) supported the necessity of metacognition as a tool for learning new skills. Conducting studies of students' abilities to monitor their own knowledge and how this relates to subsequent GPA, the two researchers found that achievement, especially in English, was better predicted by knowledge monitoring ability than raw scores in the subject of interest. This is to say that the relationship of test A to test B, given an

interval, was moderated by the variable of knowledge monitoring ability (i.e., Everson & Tobias' measure of metacognition). People with high knowledge monitoring ability increased scores significantly more from test A to test B than did those low in knowledge monitoring ability. Everson and Tobias also found that knowledge monitoring ability scores separated the students with high GPAs from the students with low GPAs. This gives support for the idea that the more skilled individuals are, the more likely they are good at gauging their own abilities.

An interesting study conducted by Hodges, Regehr, and Martin (2001) investigated the viewing of professional physician performance and its effect on novice physicians' ability to self-assess. The novice physicians performed, then rated themselves on job relevant tasks. The novice physicians were rated by experts and categorized into three ranked groups. The self-evaluations of all three groups were significantly different from the expert ratings. After viewing a benchmark comparison physician, the top group corrected their scores toward the scores of the experts, while the bottom group showed erratic corrections to their scores. Once again, higher skill was a better predictor of self-assessment ability. This also shows that self-assessment ability may be trainable in itself as a skill.

In a somewhat contradictory study, Moreland, Miller, and Laucka (1981) discovered that low-achieving students may have the capability to follow grading criteria in an objective manner, but do not apply the skill when grading themselves. That students may grade themselves less accurately than others seconds the idea of self-serving assessment ability proposed by Dunning et al. (1989). Moreland et al. (1981) instructed students to grade their own work and the work of their peers on two tests,

including a midterm and a final examination, and to describe the instructor's grading criteria. As predicted, low-achieving students did not grade their own work with accuracy, while high-achieving students did so. What was unexpected was that poor students graded their peers accurately and demonstrated an understanding of the criteria by which the instructor was grading. The authors indicated that the discrepancy could have been due to a self-serving bias, such as attributing one's performance to external and unstable factors, while attributing others' performance to stable internal factors. The student may not even consider grading criteria when grading their own responses, but use them accurately for a peer. Due to an external attributional style, a poor student may not attempt to rate one's own performance without being prompted to do so, because he/she thinks external factors are to blame. If the criteria with which the trainees are grading themselves are specific and objective, there may be less room for these biases to take effect. Also, training in a given skill, as well as training in self-assessment, should increase the skill level of an individual, allowing for more objectivity and less bias in self-assessment.

In this review, several important implications for self-assessment have been revealed. Self-assessment has been shown in several different lights with respect to its accuracy. Some research is available either to support that self-assessment is accurate (e.g., Matthews & Beal, 2002; Sullivan & Hall, 1997) or that it underestimates skill or performance (e.g., Chur-Hansen, 2000; McKinstry et al., 2003). The most common finding, however, is that self-assessment overestimates skill or performance (e.g., Barnsley et al., 2004; Dunning et al., 2004). Ambiguity was seen to have an effect on the accuracy of self-assessments, with more ambiguous criteria leading to inaccuracy and

specific criteria leading to accuracy (e.g., Ackerman et al., 2002; Oskamp, 1965).

Finally, skill level was seen to affect accuracy as well. The more skilled a person was,

the more accurately they self-assessed (e.g., Davis et al., 2006; Zakay & Glicksohn,

1992). The relationship between self-assessments, ambiguity, and skill level will be

examined further in the current study.

Current Study

The current study examined the moderating effects of criteria ambiguity and skill level on the inflation of self-assessment of skill. Although it has been shown in the literature that criteria ambiguity and skill level may affect the accuracy of self-assessment, more research is needed. The current study addressed possible interactions between criteria ambiguity and skill level and to what extent these trends can be generalized.

The current study was designed from an applied perspective with the goal of determining whether self-assessment can be used effectively by the United States Army. The military domain measuring tactical thinking skills is very specific and does not fall into the domains previously shown (i.e., Ackerman et al., 2002) to have a correlation between accuracy and the use of specific criteria and absolute scales. Furthermore, the examinations of both ambiguity and skill level, as well as the effect of ambiguity on the relationship between skill level and overestimation of skill during self-assessment, have not been examined. An individual high in skill likely is more accurate in self-impression than an individual low in skill, but they should not differ with respect to accuracy when the domain is behaviorally broken down and compared to a set of clearly defined standards. Inaccuracy is typically in the form of overestimation of actual ability. Therefore, the less skilled an individual is, the more likely he/she is to record inflated self-assessments. The introduction of criteria with more specific properties should increase the accuracy of the self-assessment and reduce inflation for unskilled individuals. Three hypotheses were tested:

H_o1: There will be a main effect of skill level on accuracy, such that Captains (i.e., higher skill level) will be more accurate than Lieutenants (i.e., lower skill level).

H_o2: There will be a main effect of item ambiguity on accuracy, such that items with greater specificity will yield more accurate self-assessments than will ambiguous items.

H_o3: An interaction will exist such that when criteria are specific, there will be no difference between Captains and Lieutenants with regard to accuracy, but when criteria are ambiguous, Captains will be more accurate than Lieutenants.

Method

The current study utilized self-assessments made by students using the Think Like a Commander (TLAC) training program. The TLAC training program (Shadrick & Lussier, 2002) is a tactical adaptive thinking training program utilized in the Armor Captain's Career Course (currently called Maneuver Captain's Career Course) at Fort Knox, Kentucky. The TLAC program is made up of seven tactical vignettes in which students are asked to identify critical information that is relevant to each vignette. The TLAC training for tactical adaptive thinking was based on an explicit set of tactical thinking behaviors identified in interviews and research with acknowledged tactical experts. The interviews and research were conducted by the Army Research Institute for the Behavioral and Social Sciences (ARI; Shadrick & Lussier, 2002).   The explicit set of tactical thinking behaviors was divided into eight themes (See Appendix A):

1. Keep a focus on the mission and higher's intent

2. Model a thinking enemy

3. Consider effects of terrain

4. Use all assets available

5. Consider timing

6. See the big picture

7. Visualize the battlefield

8. Consider contingencies and remain flexible

The criteria in the current study were extracted from these eight themes, with varying degrees of specificity.

Participants

Participants were from an active deployable unit. In this study, 22 First and Second Lieutenants and 20 Captains participated, making up a total of 42 participants. All participants were male. Both Lieutenants and Captains are Commissioned Officers in the Army, with Captains having the higher rank and more experience. To enter the Army as a Second Lieutenant, the lowest ranking Commissioned Officer, one must complete either a Reserve Officers' Training Corps (ROTC) program while earning a college degree, Officer Candidate School (OCS), U.S. Military Academy at West Point, or receive direct commission from a professional career field (United States Army, 2009a). Lieutenants and Captains completed questionnaires following a training vignette contained in the TLAC training program designed by Army Research Institute as a training aid for use during the Armor Captain's Career Course. A description and script of a training vignette in the TLAC training program is included in Appendix B. This study was reviewed and approved by the Human Subjects Review Board at Western Kentucky University (See Appendix D).

Instrument

The instrument (See Appendix C) included three parts. The first part was to complete the data concerning branch, rank, years in service, and deployment experience. Rank was used as a proxy for skill level. The second part involved a written response to the prompt, "identify and describe critical issues that should be considered in the previous vignette. Please provide reasons for your answers and make sure you address each of the relevant themes." The third part of the process was the self-assessment questionnaire. The self-assessment questionnaire consisted of 25 items designed for self-assessment of

performance during a training session (See Appendix C). The questionnaire data were archival and were collected by Dr. Scott Shadrick of ARI, Fort Knox during administration of TLAC. The instrument was developed from the Captains in Command ARI product based on the TLAC training program (S. B. Shadrick, personal communication, February 29, 2008). Objectives and self-graded items were selected from the Azerbaijan Scenario Based Vignettes to represent both low and high ambiguity items.

The first item was most ambiguous, as it evaluated the "ability to apply tactical thinking skills to the vignette" (the overall skill to be developed through the TLAC training program). The next eight items assessed the ability to perform each of the eight themes identified in the TLAC training program. These eight themes were components of tactical thinking skills and, as such, were more specific in application than the overall tactical thinking skill item. An example is, "Apply tactical thinking skills to the vignette." The first nine items (overall tactical thinking skills and eight themes) were rated on a five point Likert-type scale ranging from "no ability" to "expert ability," and reflected self-evaluations of ability to perform the given skills (See Appendix C).

The final sixteen items were specific items that applied the eight themes to the vignette in question. Each of the eight themes had two operationally defined items that comprised these sixteen items. An example of one of the sixteen questions is "Can I change my current course of action and still meet the commander's intent?" The sixteen operationally defined application items were rated either yes or no, according to self-evaluation of whether specific considerations were made during the vignette (See Appendix C).

Results

The data were analyzed by means of a 2 x 2 mixed model analysis of variance (ANOVA). The two independent variables (skill level and item ambiguity) each had two levels: high versus low ambiguity and high skill level (i.e., Captain) versus low skill level (i.e., Lieutenant). The item ambiguity variable was within-subjects, while skill level was the between subjects variable. The dependent variable was the accuracy of participants' self-assessments. The first nine items of the questionnaire served as the high ambiguity items; the final sixteen the items were low ambiguity items. The difference scores for the respective levels of item ambiguity were determined by subtracting the expert rating score from the self-assessment score. Thus, if the expert rating was higher than the self-assessment, the result was a negative number, indicating that the participant underestimated his/her ability. If the expert rating was lower than the self-assessment, the result was a positive number, indicating an overestimation of ability.

Expert ratings were used to estimate a "true score." The experts read the individual participant responses and tallied "yes" or "no" which of the 16 objective criteria were met. The expert rating was totaled for each participant by summing the number of criteria that were met. The 16 criteria were the same as the 16 low ambiguity criteria used by the individuals participating in the TLAC training program. This total represented the participants' estimated true score.

Because the high ambiguity items utilized a Likert-type scale and the low ambiguity items utilized a dichotomous "yes" or "no" scale, responses were transformed into standardized z-scores for analyses. All high ambiguity item scores were summed for each participant; this score, out of 45 possible points, was the high ambiguity self-

assessment score to undergo z-transformation. Each of the low ambiguity items was scored as either "1" for yes or "0" for no. These items were summed for a total of sixteen possible points. This served as the low ambiguity self-assessment score to undergo z-transformation. Accuracy was determined via difference scores between expert ratings of written responses to the vignette and participant self-assessment scores.

Difference scores were the best measure of accuracy in this situation. Because accuracy of assessment is the necessary dependent variable, there must be a measure of a true score for comparison. The comparison in this case is the difference between expert- and self-assessment. Some may express concern over the methodology of utilizing difference scores, doubting the reliability of the dependent variable (e.g., Kazuo & Hittner, 2003; Peter, Churchill Jr., & Brown, 1993). The argument is that the reliability of a difference score is less than the average reliability of its component variables, especially given that the component variables are positively correlated (Tisak & Smith, 1994). However, as Tisak and Smith (1994) pointed out, reliability flaws have been demonstrated primarily when a single item is used from a single source over a period of at least two points in a longitudinal study. Tisak and Smith argued that difference scores should be distinguished from change scores. Change scores refer to the measuring of one construct over time, while difference scores refer to measuring the difference between related constructs. The two related constructs in this instance involve the expert- and self-assessments. The reliability concerns that have been found with single item, single source, longitudinal data (change scores) "are not as relevant to multiple item, multiple source, cross-sectional data (p. 677)." Difference scores are not necessarily unreliable. Data from separate sources measured at the same time period are less likely to be

positively correlated than data from the same source over two time periods such that difference scores will be more reliable when using expert ratings in addition to participant self-assessments. While the concept of difference scores may be the topic of some controversy, they are the best and most feasible measure of accuracy for the given situation. Given that no completely objective measure of performance is available for the TLAC training program, the difference between the expert- and self-assessments should give the best indication of accuracy for each condition.

Tables 1 and 2 contain the results of the ANOVA. Table 1 shows the test of within subjects effects, and Table 2 shows the test of between subjects effects.

Table 1.

Test of Within Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Ambiguity	.008	1	.008	.027	.871	.001
Ambiguity * Rank	3.606	1	3.606	11.810	.001	.228
Error (ambiguity)	12.213	40	.305			

Table 2.

Test of Between Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	.057	1	.057	.034	.855	.001
Rank	25.128	1	25.128	14.914	.000	.272
Error	67.395	40	1.685			

The mixed model ANOVA yielded a significant main effect of skill level on accuracy, $F(1, 40) = 14.91$, $p < .001$, $\eta^2 = .272$. Thus, H_o1, which stated there will be a main effect of skill level on accuracy such that Captains (i.e., higher skill level) will be more accurate than Lieutenants (i.e., lower skill level), was supported. Captains in both item ambiguity conditions ($M = -.57$, SD = .76) tended to produce negative accuracy scores, while Lieutenants in both item ambiguity conditions ($M = .52$, $SD = 1.04$) produced positive accuracy scores.

There was no main effect of item ambiguity on accuracy, $F(1, 40) = 0.027$, $p = .871$, $\eta^2 = .001$. Thus H_o2, which stated there will be a main effect of item ambiguity on accuracy such that items with greater specificity will yield more accurate self-assessments than will ambiguous items, was not supported.

An interaction occurred between the two independent variables, skill level and item ambiguity, $F(1, 40) = 11.81$, $p = .001$, $\eta^2 = .228$. In the low ambiguity condition, Lieutenants overestimated ability ($M = .72$, $SD = 1.23$) and Captains underestimated ability ($M = -.79$, $SD = .82$). In the high ambiguity condition, both Lieutenants ($M = .32$, $SD = .96$) and Captains ($M = -.35$, $SD = .9$) tended toward accuracy. Thus, H_o3, which stated an interaction would occur in the opposite direction, was not supported. A graph of the estimated marginal means displays this effect in Figure 1.

Figure 1. Accuracy Interaction Between Skill Level and Item Ambiguity
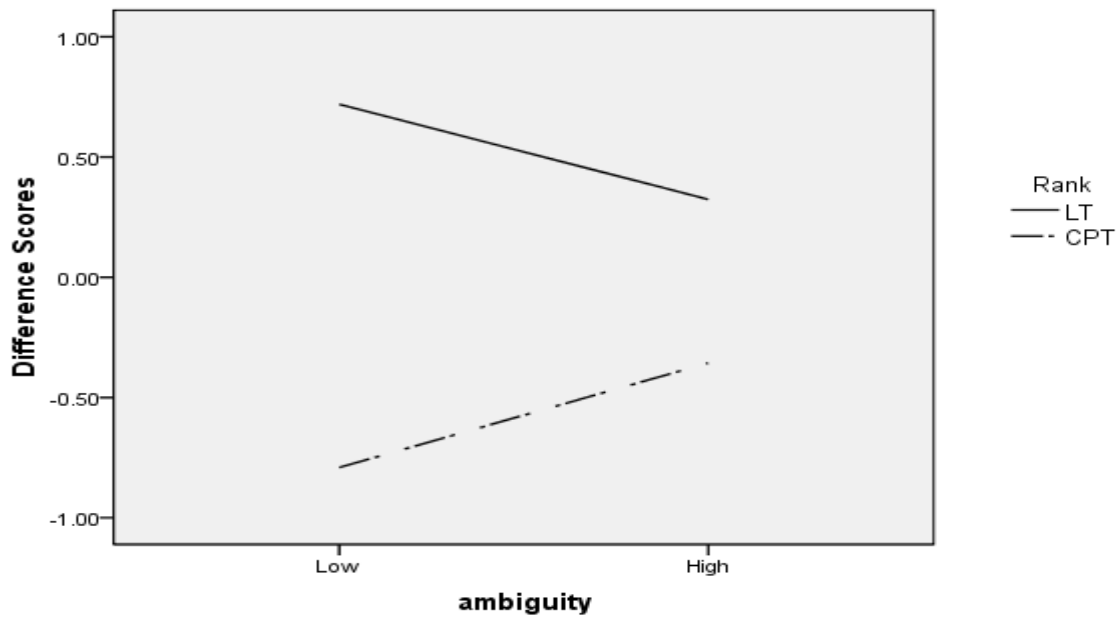
Additional Analysis

Because the expert scores for both Captains and Lieutenants were on the same scale for the less ambiguous items, they could be directly compared on the raw scores by a series of t-tests. When the difference scores on less ambiguous items for both Lieutenants ($M = 5.18$, $SD = 3.09$) and Captains ($M = 1.8$, $SD = 1.96$) are compared with a critical score of zero using a one sample t-test, both Lieutenants (t (21) = 7.85, $p < .001$) and Captains (t (19) = 4.1, $p = .001$) overestimated their scores. An independent samples t-test was conducted to examine whether the Lieutenants overestimated their performance significantly more than Captains. The groups were significantly different, t (40) = 4.18, $p = .007$, indicating Lieutenants overestimate their performance significantly more than Captains.

Discussion

Only one of the proposed hypotheses was fully supported, H_o1: There will be a main effect of skill level on accuracy, such that Captains (i.e., higher skill level) will be more accurate than Lieutenants (i.e., lower skill level). Results showed that Captains were less likely to overestimate their abilities on both specific and ambiguous items. The implications of this result include that differential considerations should be made regarding self-assessments depending on the skill level of the participant. The more skilled participant is better at self-assessing true scores of ability.

The second hypothesis was that there would be a main effect of item ambiguity on accuracy, such that items with greater specificity would yield more accurate self-assessments than would ambiguous items. The main effect of item ambiguity was not significant, so this hypothesis was not supported by the data. The interaction hypothesis, which stated that Captains would be accurate regardless of item ambiguity while Lieutenants would underestimate when items were more ambiguous and be accurate when items were less ambiguous, was not supported. An interaction was found, but it was not in the predicted direction. Instead, the interaction suggested that the more ambiguous the items, the more accurate the self-assessments. This effect is counter to prediction and counter to the current literature on item ambiguity and accuracy.

The majority of research to date reports self-assessment is inaccurate, as it typically overestimates actual ability. The results from the current study, which utilized standardized z-scores, make an interpretation of actual accuracy difficult. The raw scores were on different scales, which necessitated the use of z-scores. If all scales were on the same metric, there would be no need to utilize z-scores, and difference scores could be

30

interpreted as directly representing over- or under-estimation. By transforming the raw data into z-scores, the scores effectively went from a criterion-referenced metric to a norm-referenced metric. The use of z-scores is useful as a norm-referenced statistic, and as such is useful for investigating the differences between groups. However, without a criterion-referenced metric, it is not possible to determine if ability was over or under estimated.

Because the expert scores for both Captains and Lieutenants were on the same scale for the less ambiguous items, they could be directly compared. When the less ambiguous difference scores for Lieutenants with Captains were compared, results showed that neither Lieutenants nor Captains actually underestimated their scores for these items. For those items, the Lieutenants overestimated their performance significantly more than did Captains. The negative scores reported for Captains in both ambiguity conditions should be interpreted with caution concerning over- or under-estimation. The z-scores place the statistic on a norm-referenced scale, only using the other scores for comparison, not the original criterion.

The hypothesized effect of skill level on accuracy was supported. As an individual demonstrates more skill or aptitude for a certain behavior or domain, that individual is likely to be a more accurate self-assessor on that domain. The inference is that skilled individuals have more or better metacognitive abilities than do less skilled individuals. Compared with a highly skilled person, a less skilled person has less actual competence, but as much confidence in the self. Without high competence, the less skilled individual may allow self-serving biases to inhibit the insight necessary to realize areas of weakness. Accurately recognizing areas of weakness could serve as a motivating

force for further training. That Captains demonstrated stronger ability to self-assess suggests that self-assessment ability is something that can be developed.

The second hypothesis was not supported. H_o2 proposed that items with greater specificity would yield more accurate self-assessments than would ambiguous items. The basis for this hypothesis was research (e.g., Ackerman et al., 2002; Dunning et al., 1989; Metcalfe, 1998) stating that the lower the ambiguity, the more likely a participant or trainee will be to give an accurate self-assessment. Familiarity with ability and domain breadth (Ackerman et al., 2002), self-serving assessments (Dunning et al., 1989), and cognitive optimism outlined by metacognitive phenomena (Kruger & Dunning, 1999; 2002) are all possible reasons this occurs. As stated in the review of the literature, ambiguity, or any other single factor, is not likely to account for the entirety of self-assessment inaccuracy. The absence of a significant effect with this sample is inconsistent with the literature and does not mean that ambiguity does not affect accuracy; just that it was not found with this instrument, with this sample, and/or with this methodology.

The third hypothesis was not supported. H_o3 proposed that for the lower ambiguity condition, both Captains and Lieutenants would be accurate, but for the higher ambiguity condition, Captains would be accurate but Lieutenants would overestimate performance. What actually happened was that for the lower ambiguity condition, Captains underestimated performance and Lieutenants overestimated performance, but for the higher ambiguity condition, both Captains and Lieutenants tended toward accuracy. This is the most surprising and interesting finding in the study. The main effect of skill level was expected. The absence of a main effect for ambiguity possibly

may be a function of factors unique to this sample. The significant interaction, however, was not expected and cannot be explained by current trends in research. The implication from this finding would be that as item ambiguity increases, accuracy increases. This statement is counter to the literature review presented in this study.

Ackerman et al. (2002) stated that when assessing one's own abilities, the likelihood of under- or over-estimation is dependent upon familiarity with the ability and how broad the domain is. If the individual is unfamiliar with the ability in question, they are likely to infer from other information how they would perform. The current results could be due to the familiarity and ability of Lieutenants and Captains with the abilities tapped in the assessment. The Lieutenants and Captains participating in this study were all from an active deployable unit, meaning they were already trained to the specifications of the U.S. Army. It is possible that the global measures of the high ambiguity items were better able to capture the participants' true skills. The specific, objective items may have been so specific that they failed to capture the entire domain of performance. The domain that was addressed with the self-assessment instrument was the ability to "Enable Humanitarian Operations." The nine items that were most ambiguous tapped into the domain. An example of this type of item is the ability to "Model a thinking enemy." Because the officers are aware of what an enemy entails and are trained to know how to model a thinking enemy, their ability to assess accurately with these items is not necessarily surprising. The sixteen items that were least ambiguous also tapped into the overall domain and were targeted at more specifically measuring the abilities previously addressed by the more ambiguous items. An example of one of the specific items asks whether the participant considered "Are there other uncommitted

assets, internal to my team or at Task Force, which could provide some relief for these people?" This item demonstrates the extremely specific targeted behavior that is supposed to tap into the overall "Enable Humanitarian Operations" domain. The problem is that the targeted behavior is one of several behaviors that could demonstrate the ability to enable humanitarian operations. On one hand, the domain is very broad, but specific behavioral criteria may be necessary to objectively rate performance. On the other hand, the specific nature of the items may be excluding important domain relevant information.

An analogous concept to the broad domain broken into specific objective items is illustrated with a dartboard. If a dart is thrown at a dartboard and it hits the board in any place, the objective has been met; the dart hit its target. More specifically, however, if a dart is thrown at the bulls-eye, but hits the outer part of the board, the objective has not been met. The specific objectives are harder to meet and are harder to gauge. Although it may be simple to see whether the dart stuck in the board (i.e., the broad, ambiguous item), the specific and objective items dealing with humanitarian operations were not as easy to judge. The participants were asked whether they considered the objective items when completing the vignette. They did not have a clear-cut picture of what targets they hit or missed, and may have fallen back on the biases and metacognitive shortcomings associated with self-assessment noted in the literature.

Regardless of the inability of the current study to identify specific items as an avenue for accurate self-assessment, they should still be considered appropriate for training and performance appraisal. While the specific items may not hit their target with each question, they are important for determining whether an individual can meet individual objectives. When an item is overly ambiguous, the criterion may be

contaminated by unintentional individual objectives. Because an expert rater and a participant agree that the participant has the ability to perform well in a domain does not imply that they agree upon specific objectives within that domain.

The tasks in the literature reviewed for this study were typically procedural. Procedural tasks typically produce immediate feedback because they enable an individual to witness results as they occur. The current state of literature, which indicates that accuracy increases as ambiguity decreases, deals primarily with procedural tasks and declarative knowledge and may be specific to procedural tasks. Cognitive tasks may play a different role in self-assessment because of the difficulty of defining expert performance. Cognitive task items can be broken down behaviorally, but the thinking skills necessary for adaptive thinking are likely to remain relatively ambiguous. Adaptive expertise is when acquired skills can be applied across a variety of tasks and situations (Smith, Ford, & Kozlowski, 1997). Smith et al. (1997) described adaptive expertise as the use of detailed, structured knowledge and regulatory skills that involve processes such as planning, monitoring, and self-evaluation. The problems that many individuals now face in the workplace and in the Army are cognitively complex and demanding, and require adaptive problem solving skills. These skills are necessary to solve unstructured and ill-defined problems (Goldstein & Gilliam, 1990, in Smith et al., 1997). Adapting skills to accommodate new tasks or situations requires a deeper under standing of the components of the task, executive level capabilities to identify novel situations, and knowledge of whether or not the known set of procedures should be utilized (Smith et al., 1997). The deeper level of understanding and capabilities inherent with adaptive expertise and cognitive tasks may be the reason that self-assessment of

ability in the present study followed a different pattern than that reported in previous literature regarding item ambiguity. The ambiguous items may serve as a more accurate assessment in complex cognitive tasks that require adaptive expertise.

The current study examined a very narrow population. The target population was U.S. Army officers. The populations reviewed in previous literature often involved people from the medical and educational fields. Military officers may have a different attitude toward self-assessment. This attitude recognizes the imminent importance of training and honest self-appraisal. Officers are aware that their training could impact a life or death situation in combat, whether that life is their own or their troops'. They may not be as vulnerable to the normal biases that occur in the other populations. Thus, officers may be more likely to evaluate their skills carefully even when the domain is broad and not objectively broken down. These are questions concerning a targeted population that would be better answered with further investigation.

Limitations

The primary limiting factor in the current study was the utilization of archival data. The data were collected without regard to this research effort. The officers involved were tested following training sessions that may have left them fatigued and possibly unmotivated. The scales on the assessment instrument were the biggest hurdle in trying to decipher results. Criterion-referenced data was transformed into norm-referenced z-scores for analysis. The ability to determine over- or under-estimation was therefore limited. In the future, an instrument with consistent scales across participants and domains would be useful for the subsequent analysis and interpretation of results.

Another limitation was that the expert scores, used as participant true scores, were determined using the same scale and items the participants used for the low ambiguity items. This means that the expert scores were comparable to the low ambiguity items, but not the high ambiguity items. The difference scores were created for both levels of ambiguity using the same expert score based on the low ambiguity items. This necessitated converting the raw scores to standard scores for analysis and losing the criterion referenced nature of the raw scores. The results may have been different had both levels of ambiguity been compared with expert scores that were rated using corresponding items for each level.

Using rank as a proxy for skill level was a limitation as well. The determination of actual skill level prior to analysis would have produced a more precise measure of this variable. The utilization of rank implies that all Captains are more skilled than Lieutenants, which may not be the case.

The final limitation concerns the sample. The sample only consisted of 42 participants, so power may have suffered. Also, the sample was of a very limited population of male U. S. Army officers. This could be viewed as a major limitation. However, because the objective of this study is to apply the results to the training of U.S. Army officers, the limitation is minor. The results of this study are not intended to be generalized to the general U. S. population. While a larger sample size would have increased power, the availability of Army officers is minimal due to current military operations and training demands.

Future Research

The effect of skill level on accuracy found in this study indicates that the more skilled an individual is, the less likely they are to overestimate their abilities. The trained expert is more likely to rely on performance when estimating ability rather than extraneous information. Prior to this study, it was thought that the expert was most likely weeding out irrelevant information and utilizing pertinent information to self-assess (Kruger and Dunning, 1999; 2002). This is likely still the case. Captains were less likely than Lieutenants to over-estimate ability. Although creating specific items seemed to create objectives that were too narrow, there is reason to believe that specificity deserves attention. The more skilled participants were most likely breaking down the ambiguous items into their objective parts. The objective items were most likely too narrow and did not encompass the entire domain they were trying to measure. The objective items may reflect a criterion deficiency problem, while the ambiguous items may reflect a criterion contamination problem. If more items were utilized to capture the entire targeted domain, the relationship between ambiguity and accuracy may have been dramatically different. In the review of the literature, several studies reported that self-assessment under-estimates true performance or abilities (e.g., Chur-Hansen, 2000; McKinstry, Peacock & Blaney, 2003; Milgrom et al., 1978). Thus, the current study is not the first time that such discrepancies have been seen. Future research should address the discrepancies within the field of self-assessment, paying close attention to terminology, skill level, and item ambiguity.

The following section is a proposal for the appropriate terminology and utilization of self-assessment. The proposal coincides with the majority of research considering

specific, objective criteria as the appropriate tool for accurate self-assessment. By following the guidelines, discrepancies in the literature may eventually give way to consensus.

A Proposal for Assessment Terminology

The self-assessment literature uses terminology that confuses what could be a clear picture. In order to simplify the field of self-assessment, the terminology should be narrowed to three general terms. It is appropriate here to distinguish between self-assessment, self-grading, and self-impression. Most literature reports *self-assessment* as an estimate of how skilled/competent one-self is regarding a particular skill, ability, or characteristic. Proposed are two new terms. *Self-grading* is the assessment of one's own performance according to some objective scale. Self-grading can be the grading of one's own responses to items on a test or learning check, or grading of past performance. *Self-impression* is the overall intuitive judgment of how skilled/competent an individual feels regarding the construct. The difference between self-grading and self-impression is whether the evaluation is subjective or objective. Regarding self-impression, the evaluation can be very subjective because the individual rates his/her own perceived knowledge or standing on the construct in question. These ratings are sometimes called confidence scores (Leopold et al., 2005). In contrast, the evaluation for self-grading is less subjective because the individual rates actual performance with regard to some objective measurement system based on standards of performance set by an organization. Self-grading may be the most useful form of self-assessment for new trainees because of its ability to garner accurate self-knowledge. Self-impression may be more useful for refresher training or trainees that have advanced training as well as experience in the

environment in which the training will be ultimately put to test. This is because self-impression is an instantaneous judgment, which is more practical than time-consuming task of continually evaluating a list of individual criteria. Self-impression serves as an accurate evaluation once skills and assessment have been engrained through self-grading. The final picture of self-assessment becomes a continuum with self-grading on the one end, and self-impression on the other (see Figure 2). Along the continuum may be varying levels of how specific or ambiguous, objective or subjective the self-assessment may be.

Self-Assessment

Self-Grading   ⬅——————➡   Self-Impression
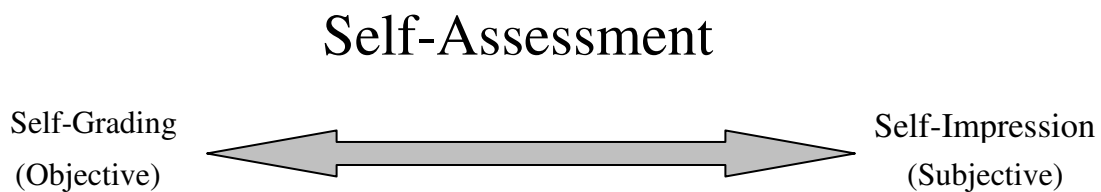(Objective)   (Subjective)

Figure 2. Self-Assessment Continuum.

Many individuals rate themselves in an inflated manner due to ambiguous standards and lack of behavioral criteria. This leads to self-impressions that are egocentric due to lack of restraint or guidelines. In order to correct for this effect of egocentric ability judgments, it is important to give the trainee an appropriate level of specificity. This will limit the ability of the trainee to exaggerate their performance. By using specific and behaviorally measurable criteria, the trainee would not be making subjective normative (compared to others) judgments, but would be making individual assessments about how well he/she performed each behavioral circumstance. After making these micro-judgments for several iterations of training, the trainee will be able to

make a blanket macro-judgment, allowing for increased breadth while maintaining accuracy. The individual criteria become engrained as a greater schema that makes up the overall assessment of performance. At this point, assessment moves from self-grading to a more global self-impression. The trainee will be able to make constant judgments of their overall performance that are indicative of training needs, providing a more practical utilization of self-assessment. The continuum between the self-grading and self-impression implies that at any time, the acuity by which one self-assesses can be adjusted according to how skilled they are. When their overall performance is deemed unsatisfactory, the trainee will be able to address the individual criteria when needed. Any problems can then be identified and corrected. Minimizing ambiguity using self-grading during training would increase feedback accuracy and eventually facilitate utilization of self-impression once mastery of a given skill has been achieved.

Skill level is another element of self-assessment accuracy. As an individual demonstrates more skill or aptitude for a certain behavior or domain, that individual is likely to be a more accurate self-assessor for that domain. The inference is that skilled individuals have metacognitive abilities that less skilled individuals do not. Compared with a highly skilled person, a less skilled person has less actual competence, but as much confidence in the self. The difference between the two is a set of biases that inhibit the insight necessary to realize areas of weakness. Areas of weakness should be recognized and seen as a motivating force for further training. If these trainees were to use specific, objective criteria to obtain a self-grade, as opposed to self-impression, their ratings would reflect more accurately their actual performance as opposed to biases that arise from lack

of competence. In this manner of self-grading, the ability to accurately rate one's self should be independent of skill level.

The individual high in skill is likely more accurate at self-impression than the individual low in skill. After training occurs, the individual that is low in skill should be both more skilled and better at self-impression. Due to the specific, behavioral nature of self-grading, its validity should be as high for the individuals with low expertise as with high expertise. The individual practicing self-grading should show high correlations with external measures of actual performance. Because skill level and self-assessment accuracy are positively correlated, we could assume that the concurrent training of skill in the domain of interest *and* self-assessment (through the recurrent use of self-grading) would be an efficient training strategy.

Similar to the argument that underlies the role of skill level is that self-assessment is a trainable skill. The literature supports a positive relationship between self-assessment accuracy and number of trials (Andrade, 2003; Edwards, 2007; MacDonald, Williams, & Rogers, 2003; Ross, 2006; Taras, 2001). That self-assessment is a skill that is transferable and independent of the domains being trained (Fitzgerald, Gruppen, & White, 2000; Schraw, 1997) gives support to the idea that care must be taken to instill this skill actively through the use of self-grading. The instinctual post-performance self-impression should be highly valid if the participant is both trained in self-assessment, and trained in the domain of interest. Self-assessment can be trained utilizing objective rating scales with behavioral measures through repeated trials of self-grading. Upon completion of training, self-assessment will be ingrained through practice of utilizing the individual
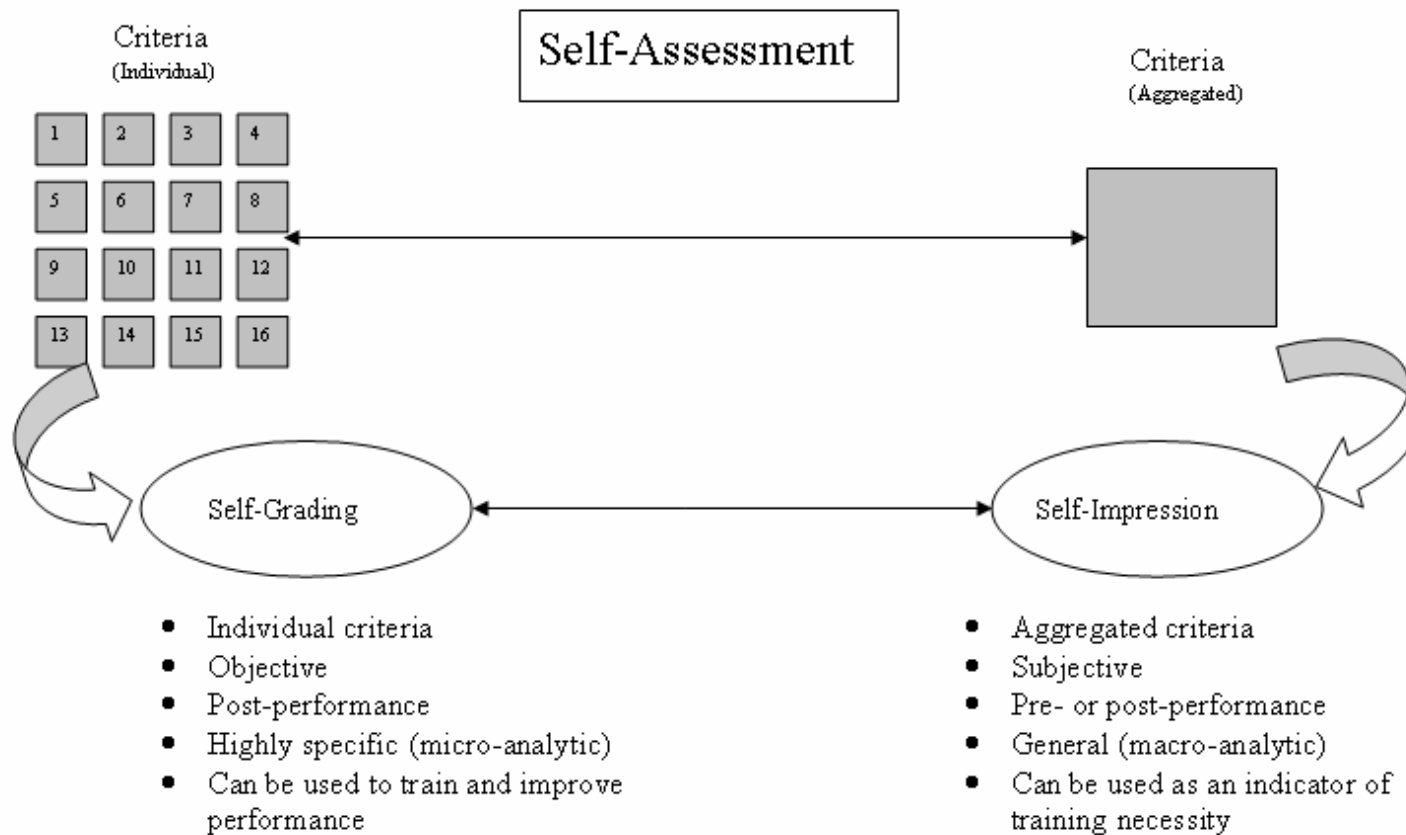
Figure 3. Self-assessment continuum contrasts the individual nature of self-grading criteria with the aggregated nature of self-impression criteria.

criteria of the self-grading process. The self-grading components can then be aggregated to fit an instinctual, instantaneous self-impression of performance (see Figure 3). In this way, self-grading of a particular skill is transformed through training into the more practical self-impression, evaluated continually during missions or future training.

Conclusion

In today's Army, the ability to self-assess is more important than ever due to increased deployment times and less time at home for schoolhouse training. In today's organizational climate, the ability to self-assess likewise is important due to flatter organizations with less time and resources for new hires and refresher training. This leads to an increased need for self-directed recognition of training needs. The accuracy of self-assessment has been underwhelming at best. The reason for this may be the misuse of self-impression when self-grading would be better suited. Although the results of this particular study point to ambiguous items as a better indication of actual performance, the abundance of self-assessment research suggests the opposite. In order to accept fully that ambiguous items are better for determining true scores, replication of the current findings is necessary. The terminology proposed in the discussion section of this paper is an attempt to clear the self-assessment waters that have been muddied by the use of similar terms for measuring self-assessments with varying degrees of specificity and objectivity.

Both military and civilian organizations should implement training that involves and utilizes the continuum of self-assessment, including self-grading and self-impression. Consideration for the situation and skill level should dictate the point within the continuum that would be most appropriate for a given training assessment. Self-grading could be useful for the introduction and training of skills that have not yet been mastered. Self-impression may also be

useful for assessing an individual's confidence or self-perception of personality or traits, as well as assessment of performance once self-assessment of a particular domain has been mastered. The implementation of the continuum has potential to improve the quality of training and skill retention throughout the organizational hierarchy.

References

Ackerman, P. L., Beier, M. E., & Bowen, K. R. (2002). What we really know about our abilities and our knowledge. *Personality and Individual Differences*, *33*, 587-605.

Andrade, H. G. (2003). Role of rubric-referenced self-assessment in learning to write. *The Journal of Educational Research*, *97*, 21-34.

Arnold, H. J. (1976). Effects of performance feedback and extrinsic reward upon high intrinsic motivation. *Organizational Behavior and Human Performance, 17*, 275-288.

Barnsley, L., Lyon, P. M., Ralston, S. J., Hibbert, E. J., Cunningham, I, Gordon, F. C., et al. (2004). Clinical skills in junior medical officers: A comparison of self-reported confidence and observed competence. *Medical Education*, *38*, 358-367.

Carless, S. A., & Roberts-Thompson, G. P. (2001). Self-ratings in training programs: An examination of level of performance and the effects of feedback. *International Journal of Selection and Assessment*, *9*, 217-225.

Castle, N., Garton, H., & Kenward, G. (2007). Confidence vs. competence: Basic life support skills of health professionals. *British Journal of Nursing*, *16*, 664-666.

Chur-Hansen, A. (2000). Medical students' essay-writing skills: Criteria-based self- and tutor-evaluation and the role of language background. *Medical Education*, *34*, 194-198.

Davis, D. A., Mazmanian, P. E., Fordis, M., Harrison, R. V., Thorpe, K. E., & Perrier, L. (2006). Accuracy of physician self-assessment compared with observed measures of competence. *Journal of the American Medical Association, 296,* 1094-1102.

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest, 5,* 69-106.

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The
    role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of*
    Personality and Social Psychology, 57, 1082-1090.

Edwards, N. M. (2007). Student self grading in social statistics. *College Teaching*, *55*, 72-76.

Everson, H. T., & Tobias, S. (1998). The ability to estimate knowledge and performance in
    college: A metacognitive analysis. *Instructional Science*, *26,* 65-79.

Fitzgerald, J. T., Gruppen, L. D., & White, C. B. (2000). The influence of task formats on the
    accuracy of medical students' self-assessments. *Academic Medicine*, *75,* 737-741.

Fox, S., & Dinur, Y. (1988). Validity of self-assessment: A field evaluation. *Personnel*
    Psychology, *41*, 581-592.

Hayes, A. F., & Dunning, D. (1997). Construal processes and trait ambiguity: Implications for
    self-peer agreement in personality judgment. *Journal of Personality and Social*
    Psychology, *72*, 664-677.

Hodges, B., Regehr, G., & Martin, D. (2001). Difficulties in recognizing one's own
    incompetence: Novice physicians who are unskilled and unaware of it. *Academic*
    Medicine, 76, 87-89.

Kazuo, K. M., & Hittner, J. B. (2003). On the relation between power and reliability of
    difference scores. *Perceptual & Motor Skills, 97*, 905-908.

Koka, A., & Hein, V. (2003). Perceptions of teacher's feedback and learning environment as
    predictors of intrinsic motivation in physical education. *Psychology of Sport and*
    Exercise, 4, 333-346.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77,* 1121-1134.

Kruger, J., & Dunning, D. (2002). Unskilled and unaware–but why? A reply to Krueger and Mueller (2002). *Journal of Personality and Social Psychology*, *82*, 189-192.

Krueger, J., & Mueller, R. A. (2002). Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance. *Journal of Personality and Social Psychology*, *82*, 180-188.

Langendyk, V. (2006). Not knowing that they do not know: Self-assessment accuracy of third year medical students. *Medical Education, 40*, 173-179.

Leopold, S. S., Morgan, H. D., Kadel, N. J., Gardner, G. C., Shaad, D. C., & Wolf, F. M. (2005). Impact of educational intervention on confidence and competence in the performance of a simple surgical task. *The Journal of Bone and Joint Surgery, 87,* 1031-1037.

Locke, E. A., & Latham, G. P. (1990). Goals and feedback (knowledge of results). In *A Theory of Goal Setting & Task Performance* (pp. 173-205). Englewood Cliffs, NJ: Prentice-Hall.

Mattheos, N., Nattestad, A., Falk-Nilsson, E., & Attstrom, R. (2004). The interactive examination: Assessing students' self-assessment ability. *Medical Education*, *38*, 378-389.

Matthews, M. D., & Beal, S. A. (2002). Assessing situation awareness in field training exercises. (Research Report 1795). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

MacDonald, J., Williams, R. G., & Rogers, D. A. (2003). Self-assessment in simulation-based surgical skills training. *The American Journal of Surgery*, *185*, 319-322.

McKinstry, B., Peacock, H., & Blaney, D. (2003). Can trainers accurately assess their training skills using a detailed checklist? A questionnaire-based comparison of trainer self-assessment and registrar assessment of trainers' learning needs. Education for Primary Care, 14, 426-430.

Metcalfe, J. (1998). Cognitive optimism: Self-deception or memory-based processing heuristics? Personality and Social Psychology Review, 2, 100-110.

Milgrom, P., Weinstein, P., Ratener, P., Read, W. A., & Morrison, K. (1978). Dental examinations for quality control: Peer review versus self-assessment. *American Journal of Public Health*, *68*, 394-401.

Mirabella, A., & Love, J. F. (1998). *Self-assessment based mini-after action review (SAMAAR) methodology: Developmental application to division artillery staff training.* (Technical Report 1086). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Moreland, R., Miller, J., & Laucka, F. (1981). Academic achievement and self-evaluations of academic performance. *Journal of Educational Psychology*, *73*, 335-344.

Oskamp, S. (1965). Overconfidence in case-study judgments. *The Journal of Consulting Psychology, 29,* 261-265.

Parker, R. W., Alford, C., & Passmore, C. (2004). Can family medicine residents predict their performance on the in-training examination? *Family Medicine, 36,* 705-709.

Peter, J. P., Churchill Jr., G. A., & Brown, T. J. (1993). Caution in the use of difference scores in consumer research. *Journal of Consumer Research, 19*, 655-662.

Randal, R., Ferguson, E., & Patterson, F. (2000). Self-assessment accuracy and assessment

centre decisions. *Journal of Occupational and Organizational Psychology*, *73*, 443-459.

Reider, B. J. (2008). *Army Self-Development Handbook*. Washington, DC: Office of the

Secretary of the Army.

Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment,*

Research & Evaluation, *11*, 1-13.

Schraw, G. (1997). The effect of generalized metacognitive knowledge on test performance and

confidence judgments. *Journal of Experimental Education*, *65*, 135-146.

Shadrick, S. B., & Lussier, J. W. (2002). Think Like a Commander: Captain's Edition –

Prototype 1.0. (Available from the U.S. Army Research Institute for the Behavioral and

Social Sciences, Fort Knox Research Unit, 121 Morande Street, Fort Knox, KY 40121-

4141).

Shadrick, S. B., & Shaefer, P. S. (2007). *Development and content validation of crisis response*

training package Red Cape: Crisis action planning and execution. (Research Report

1875). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social

Sciences.

Shoenfelt, E. L. (1996). Goal setting and feedback as a posttraining strategy to increase the

transfer of training. *Perceptual and Motor Skills, 83*, 176-178.

Sidhu, R. S., Vikis, E., Cheifetz, R., & Phang, T. (2006). Self-assessment during a 2-day

laparoscopic colectomy course: Can surgeons judge how well they are learning new

skills? *The American Journal of Surgery*, *191*, 677-681.

Smith, E. M., Ford, J. K., & Kozlowski, S. W. (1997). Building adaptive expertise: Implications

for training design strategies. In M. A. Quiñones & A. Ehrenstein (Eds.), *Training for a*

rapidly changing workplace: Applications of psychological research (pp. 89-118). Washington, DC: American Psychological Association.

Story, A. L. (2003). Similarity of trait and construal and consensus in interpersonal perception. *Journal of Experimental Social Psychology, 39*, 364-370.

Strong, B., Davis, M., & Hawks, V. (2004). Self-grading in large general education classes: A case study. *College Teaching*, *52*, 52-57.

Sullivan, K., & Hall, C. (1997). Introducing students to self-assessment. *Assessment & Evaluation in Higher Education*, *22*, 289-305.

Taras, M. (2001). The use of tutor feedback and student self-assessment in summative assessment tasks: Towards transparency for students and for tutors. *Assessment & Evaluation in Higher Education*, *26*, 605-614.

Tisak, J., & Smith, C. S. (1994). Defending and extending difference score methods. *Journal of Management, 20*, 675-682.

Ulmer, M. B. (2000, June). *Self-grading: A simple strategy for formative assessment in activity-based instruction*. Paper presented at the Conference of the American Association for Higher Education, Charlotte, NC.

United States Army. (2009a). *About the Army: Commissioned Officer*. Retrieved February 3, 2009, from http://www.goarmy.com/about/officer.jsp.

United States Army. (2009b). *Maneuver Captain's Career Course (MC3)*. Retrieved February 3, 2009, from http://www.knox.army.mil/school/16cap/mcccl.asp.

Young, J. M., Glasziou, P., & Ward, J. E. (2002). General practitioners' self ratings of skills in evidence based medicine: Validation study. *British Medical Journal, 324*, 950-951.

Zakay, D., & Glicksohn, J. (1992). Overconfidence in a multiple-choice test and its relationship

to achievement. *Psychological Record, 42*, 519-524.

Appendix A

Think Like A Commander Themes of Battlefield Thinking

Keep a Focus on the Mission and Higher's Intent.

- Commanders must never lose sight of the purpose and results they are directed to achieve—even when unusual and critical events may draw them in a different direction.

Model a Thinking Enemy

- Commanders must not forget that the adversary is a reasoning human being, intent on defeating them—it's tempting to simplify the battlefield by treating the enemy as static or simply reactive.

Consider Effects of Terrain

- Commanders must not lose sight of the operational effects of the terrain on which they must fight—every combination of terrain and weather has a significant effect on what can and should be done to accomplish the mission.

Use all Available Assets

- Commanders must not lose sight of the synergistic effects of fighting their command as a combined arms team—this includes not only all assets under their command, but also those which higher headquarters might bring to bear to assist them.

Consider Timing

- Commanders must not lose sight of the time they have available to them to get things done—a good sense of how much time it takes to accomplish various battlefield tasks and the proper use of that sense is a vital combat multiplier.

See the Big Picture

- Commanders must remain aware of what is happening around them and how it might affect their operations and how what they do can effect others' operations—a narrow focus on your own fight can get you blind-sided.

Visualize the Battlefield

- Commanders must be able to visualize a fluid and dynamic battlefield with some

  accuracy and use this visualization to their advantage—a commander who develops this

  difficult skill can reason proactively like no other.

Consider Contingencies and Remain Flexible

- Commanders must never lose sight of the old maxim that 'no plan survives the first

  shot'—flexible plans and well thought out contingencies result in rapid, effective

  responses under fire.

Appendix B

Vignette Seven Description and Script

Vignette 7: Enable Humanitarian Operations.

Mission: Your company accepts attachment of a host nation military platoon to assist with the escort of a non-governmental organizations (NGO) and private-volunteer organizations (PVO) convoy of trucks on a humanitarian aid mission. The mission is providing their escort to allow them to deliver vital supplies to a refugee camp, so speed and arriving with every truck were critical tasks in the battalion commander's intent.

Enemy: There have been occasional reports of paramilitary groups operating illegal checkpoints on the road. None of these checkpoints have been reported in the last couple of weeks.

Troops: The host nation platoon is mounted in three trucks. Your two platoons provide route recon, and close security for the trucks.

Terrain: About one-third of the way to the town, you encounter a washed out bridge. One of your platoon leaders finds a bypass that fords the river, all the trucks are able to make, and will delay your projected arrival by no more than an hour. Just as you get half way on the route, you come across a large washout jammed with trees pulled up by the roots that has choked the river almost closed. The riverbed behind is filling with water, but not very fast - the road is still high and dry and should stay so for at least a day. A few kilometers past, you crest a hill and see a large village virtually wiped out by a dam that has broken and water poured down through.

Time: Some of the NGO/PVO trucks want to stop here and help. The host nation platoon leader is contemptuous of these people and demands to continue immediately.

Civilians: You enter a town that has just been hit by flood waters. The civilians are just starting to dig themselves out and it is apparent they need help and supplies.

< Scene 1>
NARRATOR:

      The Task Force Commander has assigned your team to escort a group of trucks belonging to several non-governmental and private volunteer organizations to the Khahaly distribution center. As part of trying to reinforce the authority of the central government, a host nation platoon has been assigned to assist in the escort and security tasks. Speed is important for this convoy since the supplies are critically needed. Every truck is also vital since they have been loaded with single commodities.

< Scene 2>

      There have been occasional reports of paramilitary groups operating illegal checkpoints on the road. In the last week, no patrols have found any evidence of these checkpoints along Route Relief.

< Scene 3>
PLATOON LEADER:
"Black 6, this is White 1. The bridge up here is washed out about a kilometer past CP 37. Looks like it's pretty recent. The water is a little deep, but doesn't look too fast. I put squads out to the flanks and we found a bypass that the trucks can make. Request to take the first truck in line and send it through?"

NARRATOR:

Passage of the trucks took extra time, but worked out without any becoming stuck. Just as you get into the ceasefire boundary line area, you come across a large washout jammed with trees pulled up by the roots that has choked the river almost closed. The riverbed behind is filling with water, but not very fast - the road is still high and dry and should stay so for at least a day.

You continue moving, watching the river. The lead platoon leader (White 1) halts the convoy and calls you forward. As you crest the hill where the platoon leader stopped, you see a large town virtually wiped out by a mining catch basin that has broken and water poured through the valley.

On entering the town you see the survivors still digging out of the water, mud and ruins. They rush to the trucks, wailing and screaming, thinking the supplies are for them. Many of the things in the trucks would be useful; shovels, blankets, clothing, food, water, basic cooking items, and tents. Some of the NGO/PVO truck drivers want to stop here and help. You are very conscious of the suffering the supplies are meant to take care of in Khahlay and being behind schedule.

< Scene 4>
HOST NATION PLATOON LEADER (accented English):
"Not stopping here! No help! We GO NOW!"

VILLAGE ELDER (translated):
"He says: Praise the all Supreme One whose Name we are not fit to pronounce! Our prayers have been answered! I called for help hours ago and you have arrived! What have you brought? How many men? We can still save some of the people who were buried in the mud! This way, this way!"

NARRATOR:
An attempt to call the Task Force to alert them to the crisis here is unsuccessful due to the steeply sided, narrow valley you are in. It would take over 20 minutes to reach a position where the radio would be able to make contact.

Appendix C

Think Like A Commander Self-Assessment Instrument

Vignette 7: Enable Humanitarian Operations

Name: _____   Branch: _____

Rank: _____   Years in Service: _____

Deployment Experience (Location, Dates, and Duty Position):

_____

_____

_____

_____

_____

Please identify and describe the critcal issues that should be considered in the previous vignette.
Please provide reasons for your answers and make sure you address each of the relavent themes.

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Think Like A Commander
Vignette 7
Enable Humanitarian Operations

For the following set of questions, **refer to your written response** *to the vignette.*
Please indicate your ability to perform the indicated skill by **checking the**
appropriate box *.*

	No Ability	Little Ability	Average Ability	Satisfactory Ability	Expert Ability
1. Apply tactical thinking skills to the vignette	☐	☐	☐	☐	☐
2. Keep a focus on the mission and higher's intent	☐	☐	☐	☐	☐
3. Model a thinking enemy	☐	☐	☐	☐	☐
4. Consider effects of terrain	☐	☐	☐	☐	☐
5. Use all assets available	☐	☐	☐	☐	☐
6. Consider timing	☐	☐	☐	☐	☐
7. See the big picture	☐	☐	☐	☐	☐
8. Visualize the battlefield	☐	☐	☐	☐	☐
9. Consider contingencies and remain flexible	☐	☐	☐	☐	☐

Please indicate (yes or no) whether you considered each of the following questions
while writing your response to the vignette.

	Yes	No
1. Can I change my current course of action and still meet the commander's intent?	☐	☐
2. Was I capable of performing both tasks--escorting the convoy to Khahaly and helping the people in the village--simultaneously?	☐	☐
3. What can I expect to see downstream? Can I even reach Khahaly?	☐	☐
4. How do the situations at Khahaly and Zarysly compare?	☐	☐
5. Are there sufficient resources on hand to implement an alternate course of action to satisfactorily resolve both problems?	☐	☐
6. Are there other uncommitted assets, internal to my team or at Task Force, that could provide some relief for these people?	☐	☐
7. How can watercourses affect the rest of my planned movement to Khahaly?	☐	☐
8. What steps can be taken to prevent further damage/flooding and prepare for subsequent relief efforts?	☐	☐
9. How should I deal with the reluctant Azerbaijani platoon leader?	☐	☐
10. How could my decision lead to a perception that U.S. forces favor one group over another?	☐	☐
11. How have events already affected my original mission's timing?	☐	☐
12. How long will it take higher to get a response team into position and begin providing assistance to the village?	☐	☐
13. Do I have sufficient justification to abort or modify my mission?	☐	☐
14. How do I handle the possiblility that my actions will result in failure to meet deadlines associated with the original mission?	☐	☐
15. Can I expect to meet active resistance by hostile forces?	☐	☐
16. Can I "fight" on two fronts and still be effective?	☐	☐

Appendix D

Western Kentucky University Human Subjects Review Board Approval

WKU®

A LEADING AMERICAN UNIVERSITY WITH INTERNATIONAL REACH
HUMAN SUBJECTS REVIEW BOARD

In future correspondence, please refer to HS09-195, April 2, 2009

John Breidert
c/o Dr. Betsy Shoenfelt
Psychology
WKU

John Breidert:

Your research project, *Self Assessments by U.S. Army Officers: Effects of Skill Level and Item Ambiguity on Accuracy*, was reviewed by the HSRB and it has been determined that risks to subjects are: (1) minimized and reasonable; and that (2) research procedures are consistent with a sound research design and do not expose the subjects to unnecessary risk. Reviewers determined that: (1) benefits to subjects are considered along with the importance of the topic and that outcomes are reasonable; (2) selection of subjects is equitable; and (3) the purposes of the research and the research setting is amenable to subjects' welfare and producing desired outcomes; that indications of coercion or prejudice are absent, and that participation is clearly voluntary.

1. In addition, the IRB found that you need to orient participants as follows: (1) signed informed consent is not required; (2) Provision is made for collecting, using and storing data in a manner that protects the safety and privacy of the subjects and the confidentiality of the data. (3) Appropriate safeguards are included to protect the rights and welfare of the subjects.

This project is therefore approved at the Exempt Review Level.

2. Please note that the institution is not responsible for any actions regarding this protocol before approval. If you expand the project at a later date to use other instruments please re-apply. Copies of your request for human subjects review, your application, and this approval, are maintained in the Office of Sponsored Programs at the above address. Please report any changes to this approved protocol to this office. A Continuing Review protocol will be sent to you in the future to determine the status of the project. Also, please use the stamped form that accompanies this letter.

Sincerely,

Paul J. Mooney, M.S.T.M.
Compliance Manager
Office of Sponsored Programs
Western Kentucky University

cc: HS file number Breidert HS09-195

The Spirit Makes the Master
Office of Sponsored Programs | Western Kentucky University | 1906 College Heights Blvd. #11026 | Bowling Green, KY 42101-1026
phone: 270.745.4652 | **fax:** 270.745.4211 | **e-mail:** paul.mooney@wku.edu | **web:** http://ored.wku.edu/Research_Compliance/Human_Subjects/
Equal Education and Employment Opportunities • Printing paid from state funds, KRS 57.375, 2006 • Hearing Impaired Only: 270.745.5389